DOCUMENT RESUME

ED 159 502 CS 004 295

AUTHOR Luiten, John; And Others
TITLE Using Fry's Graph to Describe the Variation of Readability: A Corrected Procedure.
PUB DATE [78]
NOTE 14p.; Study prepared at the University of Arizona

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Basic Reading; Elementary Secondary Education; *Evaluation Methods; Predictive Measurement; *Readability; *Readability Formulas; *Reading Research; Textbook Evaluation; Validity
IDENTIFIERS *Fry Readability Graph

ABSTRACT

A method proposed by McCuaig and Hutchings to assess the internal readability variation of books using the Fry readability formula was empirically tested in an analysis of six elementary basal and ten secondary remedial readers. Each book was treated as a series of consecutive 100-word samples. Means and standard deviations were computed for both sentence length and syllable count measures. Using this information, several different pairs of intervals were constructed for each book using McCuaig's and Hutchings' method. These intervals were designed so that their intersections resulted in 11 different-sized zones, which, according to McCuaig and Hutchings, should have contained from 10 to 95% of the samples within the text. The analysis indicated that this procedure consistently underestimated the readability variation. The procedure, however, can be corrected with one alteration: the probability of a sample lying within the zone formed by the intersection of the two intervals is equal to the product of the separate probabilities of these intervals (the law of independent events). (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED159602

Using Fry's Graph to Describe the Variation of Readability: A Corrected Procedure

John Luiten

Department of Educational Psychology

University of Arizona

Wilbur Ames

Department of Reading

University of Arizona

John Bradley

Department of Reading

University of Arizona

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

John Luiten, Wilbur Ames, John Bradley

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Running head: Variation of Readability: A Corrected Procedure

CS004295

Abstract

A method proposed by McCuaig and Hutchings to assess the internal readability variation of books using the Fry Readability Formula was empirically tested. In each of sixteen books examined, the McCuaig and Hutchings method was found to consistently underestimate the observed readability variation. An explanation of this underestimation is presented. In addition, the authors of this paper propose a change in the logic of McCuaig and Hutchings' method which eliminates this underestimation.

Using Fry's Graph to Describe the Variation of Readability: A Corrected Procedure

To date, little information concerning the readability of elementary and secondary texts has been made available by publishers. Aside from a book's grade designation, little else is typically provided about its readability properties. While information concerning a book's average readability in the form of a grade designation has some value, it can also be misleading by serving to mask the extreme readability variation likely to be present. Based on a variety of instructional materials, there is evidence now available which shows that considerable intrabook readability variation can exist (1, 2, 4) and that it does affect performance (3).

The importance of assessing the variation of difficulty within a book has been too often overlooked. In one of the few treatments of this topic, McCuaig and Hutchings (7) emphasized the importance of assessing internal readability variation in books. In addition, they presented a specific method for assessing difficulty variation through the use of the Fry Readability Graph (6). Fry's Readability Graph

employs only two variables--number of syllables per 100 words (syllable count) and number of sentences per 100 words (sentence length). These two measures, when plotted on the graph supplied by Fry, yield an estimate of reading difficulty in terms of grade level.

McCuaig and Hutchings have suggested the following procedure to estimate internal readability variation of a book. First, they advised drawing six 100-word samples evenly distributed throughout the book for use to make counts of the sentences and syllables according to the Fry procedure. The mean and standard deviation for the syllable count and sentence length measures are then computed. Next, using one standard deviation as a distance above and below the mean, they suggested creating an interval about each of the obtained means. The intersection of these two intervals on the Fry Graph is then, determined. This intersection creates a zone which can be shaded on the Fry Graph, and which McCuaig and Hutchings felt adequately describes readability variation in the measured book (Figure 1). Specifically, they stated that this zone of plus and minus one standard deviation about the mean in size would contain 68 percent of all 100-word samples within the book.

Figure 1 here

McCuaig and Hutchings suggest that a zone containing any desired percentage of running text within a book may be constructed by changing the size of the intersecting intervals. One can then assess the

variation in readability within the book by determining the grade levels encompassed by the shaded zone on the Fry Graph. For example, suppose that the shaded zone in Figure 1 had been constructed by using intersecting intervals each of which extended from two standard deviations below the mean to two standard deviations above the mean. According to McCuaig and Hutchings, a zone of that size (± 2 S.D.) should contain 95 percent of the running text within a book. Using this information, a teacher could say that almost all of the text within the book is between the sixth and ninth grade level in readability and further, that most of the text is at a seventh grade level, less at the eighth and ninth grade levels, with hardly any at the sixth grade level. However, the usefulness of this procedure depends upon the degree to which the zone contains the estimated (predicted) number of samples. Therefore, it was the purpose of this study to empirically test the accuracy of McCuaig and Hutchings' procedure for predicting the percent of text in zones of differing sizes on the Fry Readability Graph.

Sixteen books, six elementary basal and ten secondary remedial readers, were chosen for analysis (Table 1). Following the procedure suggested by Clymer (5), each book was treated as a series of consecutive 100-word samples.

Beginning with the first word of the first story, one hundred words were counted according to Fry's rules, and sentence length and syllable counts were made. This procedure was followed for each succeeding 100-word passage to the end of the story. At the end of

a story, there were always less than one hundred words left. Rather than have a sample consist of parts from two different stories, any 'left-over' words were ignored and the next sample count was begun with the first word of the next story. This procedure was followed separately until the end of each book yielding 'book' counts for syllables and sentences. The shortest book had 91 samples while the longest had 843. For each book, means and standard deviations were computed for both the sentence length and syllable count measures. These two statistics, mean and standard deviation, were considered to be the parameter values for each book.

Using the obtained means and standard deviations, several different pairs of intervals were constructed for each book using McCuaig and Hutchings' procedure. These intervals were designed so that their intersections resulted in eleven different-sized zones which, according to the logic of McCuaig and Hutchings' procedure, should have contained from 10 to 95 percent of the samples within the text. Finally, all samples for each of the sixteen books were plotted on the Fry Graph to determine what percentage of text in each book actually fell within each of the different-sized zones. Table 1 shows the results obtained with the above procedure for the sixteen books analyzed.

Table 1 here

For all books analyzed, there was a consistent tendency for McCuaig and Hutchings' procedure to overestimate the percentage of text found within any of the eleven zones. This bias is shown by examining the results obtained with individual books as well as the mean scores across all sixteen books. The overestimation bias was most pronounced with the smaller zones and decreased as zone size increased. It is interesting to note that McCuaig and Hutchings' procedure was not accurate for even one book or one zone.

McCuaig and Hutchings' procedure involves two assumptions. One of these they have termed "limited independence" and the other simply "normality". The first assumption involves the correlation between the syllable count and sentence length measures while the second refers to the distribution of these two measures.

Exactly what is meant by limited independence is never fully explained by McCuaig and Hutchings, but can be surmised by the implied logic of their procedure. In their original example, an interval of one standard deviation about the mean has a probabiltiy density of .68 (assuming normality). If the intersection of the two intervals, the zone, is to contain 68 percent of the samples, then limited independence must be understood to mean <u>total dependence</u> or <u>perfect correlation</u> between the syllable count and sentence length measures. Any departure from this relationship will cause fewer than 68 percent of the text to fall within the zone.

Assuming the above to be true, and further, assuming that the

sentence length and syllable count measures are indeed normally distributed, then one should be able to account for the bias of McCuaig and Hutchings' procedure by examining the correlation between these two measures.

The Pearson correlation coefficient between the syllable count and sentence length measures for each book analyzed as well as the mean correlation coefficient (obtained by Fisher's Z method) for all sixteen books are presented in Table 1. The magnitude of the coefficients ranged from a high of -.55 to a low of -.02 with a mean of -.13. These data are more supportive of a complete independence assumption than a complete dependence assumption.

Independence being the case, McCuaig and Hutchings' procedure can still be used with one minor alteration--the probability of a sample lying within the zone formed by the intersection of the two intervals is equal to the product of the separate probabilities of those intervals (law of independent events). In the example presented by McCuaig and Hutchings, therefore, the correct probability of a sample lying within the zone is .68 x .68, or .46. One can expect 46 percent (.46 x 100) of the samples in the book to fall within the zone, not 68 percent as originally claimed by McCuaig and Hutchings.

The last line of Table 1 shows the uncorrected McCuaig and Hutchings' prediction for each zone. Directly above these (second line from bottom), is the corrected prediction for each zone. A simple comparison of the uncorrected predictions and the corrected pre-

dictions with the obtained mean results for the sixteen books analyzed clearly shows the corrected predictions (using an independent assumption) to be superior to the uncorrected predictions (using a dependence assumption) in estimating the percentage of text found in each zone. Because of the superior fit of the corrected prediction to the results obtained, non-normality of the syllable count and sentence length measures is not considered to be a factor affecting prediction.

It appears that McCuaig and Hutchings' procedure as originally proposed does not yield an accurate assessment of internal variation in book readability. This inaccuracy stems from predictions which consistently overestimate the percentage of text found in any zone constructed. The correct procedure for estimating the percentage of samples within a zone has, therefore, been presented in this paper. The reader is reminded, however, that this study used the population mean and standard deviation for each book in the construction of the different-sized zones. McCuaig and Hutchings suggest using a random sample of six 100-word passages to estimate these parameter values.

It is suggested that research be conducted to determine the sample size needed to adequately estimate a book's readability variation when the corrected McCuaig and Hutchings procedure is employed.

REFERENCES

1. Ames, Wilbur S. and John M. Bradley. "Readability Parameters of Secondary Remedial High-Interest Materials," unpublished manuscript, 1978.

2. Ames, Wilbur S. and John M. Bradley. "An In Depth Study of the Readability of a Social Studies Textbook," unpublished manuscript, 1978.

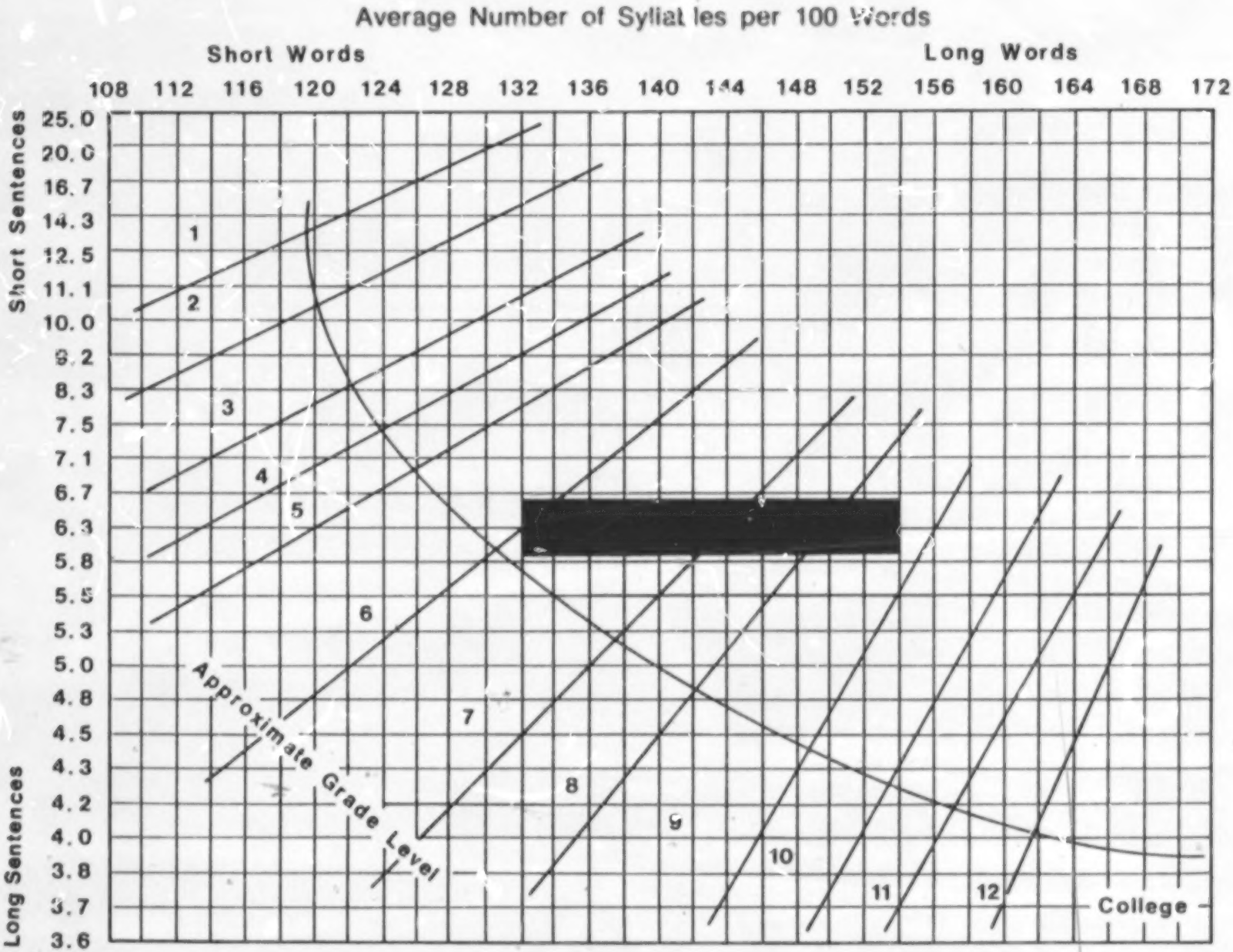
3. Bradley, John M. and Wilbur S. Ames. "The Influence of Intrabook Readability Variation on Oral Reading Performance," Journal of Educational Research, 70: 101-105, 1976.

4. Bradley, John M. and Wilbur S. Ames. "Readability Parameters of Basal Readers," Journal of Reading Behavior, 9: 175-183, 1977.

5. Clymer, Theodore. "A Study of the Sampling Reliability of the Spache Readability Formula," Reading in a Changing Society. Procedings of IRI, 4: 245-250, 1959.

6. Fry, Edward. "A Readability Formula That Saves Times," Journal of Reading, 11: 513-516, 575-578, 1968.

7. McCuaig, Susannah M. and Barton Hutchings. "Using Fry's Graph to Describe the Variation of Readability," Journal of Reading, 18: 298-300, 1975.



Graph for estimating readability, by Edward Fry, Rutgers University Reading Center.

Table 1. Percent of Text in Various Zones, Correlation between Sentence Length and Syllable Count, and Corrected Versus Uncorrected McCuaig and Hutchings' Predictions

BOOK TITLE	Percent of Text in Zones (Columns Labeled by Interval Size in Standard Deviation Units)											
	± .125	± .255	± .385	± .525	± .675	± .840	± .995	±1.040	±1.280	±1.645	±1.960	r
Secondary Remedial												
The Time is Now	1	2	5	13	31	42	51	52	70	84	88	-.12
Cities	2	3	6	11	21	32	46	47	65	82	93	-.18
Out of Sight	0	1	5	13	24	31	42	47	63	83	91	-.03
Larger Than Life	0	4	11	15	22	35	48	48	62	81	90	.14
Where Its At	1	3	7	14	21	36	44	50	68	83	90	-.16
With It	3	3	8	12	20	28	47	51	66	87	94	-.15
Beyond the Block	0	4	8	14	21	30	45	46	70	82	91	-.12
Over the Ridge	0	2	9	17	26	32	41	45	62	83	94	-.07
Communication	0	1	6	14	25	44	52	58	73	89	91	-.35
This Land	1	4	11	16	22	43	53	57	68	83	89	-.55
Elementary Basal												
Panorama	1	8	14	20	30	46	59	63	75	89	94	-.19
Secret Spaces	0	3	6	11	16	28	39	45	62	82	91	-.09
Secrets	0	2	6	11	21	35	42	47	64	83	91	.05
Never Give Up	1	3	9	18	26	35	47	52	65	82	92	.12
Good News	2	7	12	16	23	45	56	56	73	86	92	-.02
Kaleidoscope	1	4	9	19	27	39	52	53	68	84	90	-.24
MEAN	1	3	8	15	24	36	48	51	67	84	91	-.13
McCuaig & Hutchings' Prediction												
With Correction	1	4	9	16	25	36	46	49	64	81	90	
Without Correction	10	20	30	40	50	60	68	70	80	90	95	